THE OHIO STATE UNIVERSITY

An investigation of biases in Patient Safety Indicator score distribution amongst hospital cohorts

A Thesis Presented in Partial Fulfillment for Graduation with Distinction from the School of Health and Rehabilitation Sciences

Santino Cua
Health Information Management and Systems
School of Health and Rehabilitation Sciences
4/11/2016

Graduation with Distinction Committee:
Susan E. White, PhD, RHIA, CHDA, Advisor
Susan D. Moffatt-Bruce, MD, PhD, FACS, Advisor
Carmen P. Digiovine, PhD, ATP, RET

Abstract

The Centers for Medicare and Medicaid Services (CMS) have implemented a hospital reimbursement system that incentivizes payment proportional to the quality of care delivered and performance on certain metrics. One such metric is the Agency for Healthcare Research and Quality's Patient Safety Indicator 90 (PSI-90). It is composed of eight individual indicators designed to flag adverse patient events that are potentially preventable, such as post-operative wound dehiscence and accidental lacerations. CMS publicly reports four of these individual PSI scores (6, 12, 14 and 15) in addition to the composite PSI-90. Previous studies question the PSIs' validity beyond screening purposes and furthermore question the underlying administrative data's ability to accurately and reliably flag such events. This study looks to analyze biases in PSI score distribution for hospitals depending on teaching status, differences in patient demographics and lastly, interactions between teaching status and patient demographic factors and their ability to account for differences in PSI rates. Significant differences were found between teaching and non-teaching hospitals for PSIs 6, 12, 15 and 90 ($\rho \leq 0.01$). Inpatient volume and patient severity ($\rho \leq 0.01$) were found to be significantly different between teaching status cohorts. Lastly, significant differences in PSI scores were found between patient severity quartiles for PSI 6, 15 and 90 ($\rho \leq 0.05$) and between socio-economic quartiles for PSI 6, 12, 15 and 90 ($\rho \leq 0.05$); but interaction between patient severity and teaching status was only significant for PSI 90 ($\rho \leq 0.05$) and between socioeconomic and teaching statuses for PSI 6 ($\rho \leq 0.05$). These results indicate current PSI score distributions may be biased against teaching hospitals for 4 out of 5 PSI measures. Further studies will involve assessing the

adequacy of risk-adjustment methodology for PSI metrics. Until then, use of PSI metrics to determine federal reimbursement can lead to bias against teaching hospitals.

Acknowledgements

I would like to thank my advisors, Dr. Susan E. White and Dr. Susan D. Moffatt-Bruce for their time, cooperation and encouragement throughout my research process. Without their willingness to meet with me, to edit my work, and to provide me with countless resources and tips throughout the process, the completion of my undergraduate thesis would not have been possible.

List of Tables

List of Figures

Table of Contents

Introduction

Problem Statement

Policy reform has shifted the United States' healthcare focus to facilitating higher-value care by providers. The government has implemented a number of metrics to determine higher "value", or quality of care, including measuring readmission rates, patient outcomes, satisfaction and safety, among others. Subsequently, they have incentivized compliance by determining federal reimbursement proportional to metric performance. One metric of particular concern is the patient safety indicator score, or PSI. PSI scores were initially designed for screening purposes but have evolved beyond their original scope to now measure quality of care, and furthermore, to determine reimbursement. There are integral concerns about the adequacy of PSI scores in determining quality of care and thus, allocating reimbursement proportionately. This potentially introduces inefficiency into the United States' healthcare system as biased reimbursement implicates millions of dollars that could negatively affect medical decision making and provision of care.

Patient safety became a priority in the United States healthcare system after the publishing of the report, *To Err is Human*, in 2000, which estimated up to 98,000 deaths annually were due to preventable medical errors in the U.S. (Institute of Medicine, 1999). The Agency for Healthcare Research and Quality (AHRQ) had already developed quality indicators in 1994 based on the Healthcare Cost and Utilization Project. These 33 indicators included measures such as mortality rates and rates of procedures to evaluate resource utilization in healthcare. Over the years, knowledge expanded concerning evaluation of such indicators, as did the access to data used to determine them. Thus, AHRQ tasked researchers at Stanford and the University of California-San Francisco to refine the scope and sophistication of these indicators (Haytham M.A. Kaafarani & Rosen, 2009).

The collaboration resulted in 20 provider-level PSI algorithms based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnoses, procedures and diagnoses-related groups (Agency for Healthcare Research and Quality, 2008). Of the 20 provider-level PSIs, 11 are included in PSI-90, a composite score for comparison amongst hospitals. An additional 7 area-level PSIs were developed after for comparisons between populations (i.e. geographic areas or health plans). Provider-level indicators flag potentially preventable complications via secondary diagnoses and procedure code flags. Area-level indicators similarly include secondary diagnoses but are also specified to capture primary diagnoses to identify complications from another hospitalization (Agency for Healthcare Research and Quality, 2014). PSI scores are calculated by dividing the number of discharges with indication of a particular adverse event by all discharges at risk for the adverse event (Agency for Healthcare Research and Quality, 2008). These rates are then adjusted based on reference population to minimize influence of outliers. ICD-9 codes are integral to the calculation of PSIs, as the validity of the scores is directly related to coding reliability.

ICD-9 emerged as a coding classification system for hospital inpatient billing and reimbursement in 1979. Accuracy is paramount as ICD-9's applications begin to grow beyond billing purposes, and can directly influence the quality of care delivered due to decisions made relative to coded diagnoses. O'Malley et al (2005), describes coding accuracy as the ability to represent the underlying disease or condition. ICD-9 codes are now evaluated for their reliability in flagging PSIs and adverse events, or events that result in harm to a patient due to care; however, concerns exist about the classification system's validity (Haytham M.A. Kaafarani & Rosen, 2009).

While PSIs were invented for internal quality improvement purposes and identifying adverse events (H.M.A. Kaafarani, Borzecki, & Itani, 2011), PSIs are now being used for hospital profiling and federal pay for performance purposes. They currently incorporated into CMS' hospital value-based purchasing program (VBP) for inpatient stays (CMS 2013). This program was integrated in the 2010 Affordable Care Act (ACA) because of demand for a shift in US healthcare agenda-- rewarding quality of care over volume of care.

VBP was implemented at the beginning of Fiscal Year (FY) 2013, or Oct 1st, 2012, in which Medicare withheld 1% of reimbursement for all Medicare inpatient discharges to generate a pool for performance based incentive payments. In FY 2013 alone, this pool was projected to be $850 million. The percent withheld will increase by .25% annually until it is 2% by FY 2017 (Centers for Medicare & Medicaid Services, 2013). For FY 2013, incentives are only based on two domain measures: Process of Care and Patient Experience of Care. As of 2015, VBP has expanded to incorporate two more domain measures, Efficiency and Outcome. Included in the Outcome domain is the AHRQ's PSI-90 composite score. The Outcome domain is weighted to account for 30% of a hospital's total performance score- the final determinant of reimbursement level (Centers for Medicare & Medicaid Services, 2013).

Studies involving ICD-9 codes and administrative data have yet to provide conclusive results that they are reliable and accurate input sources for flagging preventable adverse events (Tsang, Palmer, Bottle, Majeed, & Aylin, 2011). Studies involving PSIs' validity beyond screening purposes such as performance measures have equally been called into question due to a lack of a reliable measure of severity and specificity of a patient's underlying conditions in billing data (Tsang, Palmer, Bottle, Majeed, & Aylin, 2011). Despite these uncertainties, the federal government is now determining hospital reimbursement based on measures that may not encompass a true representation of care received.

Objectives of the Study

The objective of the study is to analyze PSI score distribution across hospital cohorts, specifically incidence of higher rates of PSIs in teaching hospitals versus nonteaching hospitals. Teaching hospitals tend to employ higher intensity of service for their patient populations. Variables, such as inpatient volume, patient severity and patient socio-economic status will be included in the analysis to determine if they are statistically different amongst facility type. The impact of these covariates can be determined by employing VBP reimbursement metrics dependent on PSI scores. Lastly, the PSI scores can be tested for correlation with other claims-based quality metrics for the hospital cohorts. These objectives will be addressed by investigating the following research questions:

1. Are the PSI scores different in teaching versus non-teaching hospitals?
2. Are the patients treated by teaching hospitals different than those treated by non-teaching hospitals as measured by inpatient volume, patient severity and socio-economic status?
3. Is there an interaction between the teaching status of a hospital and the patient demographics that helps explain the difference in PSI scores between teaching and non-teaching hospitals?

Literature Review

Factors effecting Patient Safety Indicators

Naessens et al (2012) investigated admission health status on hospital adverse events and associated costs. The study considered the effects of comorbidities and illness severity on adverse events and PSIs among adult medical and surgical patients. They employed PSI software as the indicators are designed to flag based on secondary diagnosis codes; however, the software sometimes includes PSIs that are present on admission (POA) and not attributable to current care (Naessens et al., 2012). PSIs are programmed to adjust for patient demographics, diagnosis related groups and comorbidities but do not adjust for physiological state of the patient on admission. Therefore, Naessens et al (2012) employed acute physiological score (APS) that includes additional clinical data, lab values and physiological assessments. The results of the study showed a positive association between incidence of a PSI during stay and increased admission illness severity as measured via the APS. Subsequently, presence of certain comorbidities, rather than number of them, also linked with increased likelihood to experience a PSI during stay. Lastly, the inclusion of demographics, DRGs, POA conditions, APS score and specific comorbidities into a risk-adjustment model increased the number of adverse events indicated- suggesting an underrepresentation with the current model (Naessens et al., 2012).

Patients that are more severely ill require more complex, frequent and demanding treatment that can increase their probability of experiencing harm due to medical error. This is depicted by a positive correlation between PSI incidence during stay and an increased admission illness severity. Hospitals that admit higher volumes of sicker patients could fall victim to public reporting of decreased quality and safety, and deliver more uncompensated care given current VBP metrics (Naessens et al., 2012). While AHRQ has since incorporated exclusion of conditions POA into their PSI software, they still fail to address the patient's physiological state on admission. Consideration of APS score in PSI software would help to create a more representative measure that appropriately adjusts for hospitals that care for higher volumes of severely ill patients (Naessens et al., 2012). If not, such hospitals and systems may consider reducing admission and delivery of care to such patients in order to avoid federal reimbursement penalties.

A similar study in 2013, Rahman, Neal, Fargen and Hoh conducted multivariate analysis on the effect of gender, age, comorbidities, hospital type and size on the probability of experiencing a PSI or hospital-acquired condition (HAC), another federal quality measure. In particular, this study found that for PSIs, increasing the comorbidity score caused either linear or exponential increases in probability of all PSIs. Subsequently, the study determined that a patient admitted to a nonteaching hospital was significantly less likely to be diagnosed with a PSI compared to a teaching hospital (Rahman, Neal, Fargen & Hoh 2013). Lastly, the study deduced that increase in incidence of PSIs was positively correlated with increased length of stays, mortality rates and hospital charges. The findings potentially describe why nonteaching hospitals are associated with lower incidences of PSIs, and thus higher quality by federal standards (Rahman, Neal, Fargen & Hoh's 2013). Furthermore, teaching hospitals are forced to assume higher costs and risks by delivering care to patients likely to be diagnosed with a PSI during their stay.

Considering the demographics of the patient populations of teaching hospitals, the current VBP metrics harbor an inherent bias against teaching hospitals in reimbursement. Thus, there is a shift in focus on improving coding accuracy and pre-admission patient profiling in order to reduce the risk of patient safety events (Rahman, Neal, Fargen & Hoh 2013). Current financial penalties for high incidences of PSIs may also lead to reduced documentation and reporting. Furthermore, health systems may be unwilling to admit high-risk patients. Such repercussions would be counter to the goals of recent legislation in improving access to and quality of care.

ICD-9 Coding Accuracy

ICD-9 code assignment is a complex process as it is greatly influenced by multiple factors throughout the processes of care. In a landmark study, O'Malley et al (2005) identified common sources of error in ICD-9 coding in two domains: patient trajectory through a facility and medical record trail. A patient begins their process of care upon registration and admission to a facility. Failure to acquire quality information at intake invites error and accuracy to all subsequent processes, such as failure to record a condition POA and its effects on coding and administrative data (O'Malley et al, 2005). Additionally, error can be introduced throughout a patient's trajectory via miscommunication between clinical staff and providers, specificity in their documentation of care (O'Malley et al, 2005). Furthermore, physicians often do not document with the intent of the record being used for classification and billing purposes (Wil, 2014). Dr. Wil Lo asserts that while lack of clinical documentation training affects the integrity of the record, time to document patients with high complexity care is much more cumbersome and may lead to non-compliance and inadequacy in the medical record (2014). The process of creating and editing the medical record is distinguished as its paper trail in this study. Subsequently, coding professionals must use these potentially inadequate medical records to properly assign MS-DRG and ICD-

9 codes. However, there can be discrepancies amongst coders in their interpretation of the medical record as well (Alonso & Love, 2013). This can lead to opportunities for errors in holistic, accurate data to represent additional elements beyond basic code assignment such as PSIs, severity of illness, or hospital-acquired conditions (HACs) among others (Alonso & Love, 2013). Ultimately, these errors and inadequacies hinder reimbursement optimization and can affect medical decision-making and research studies, areas of great importance in the healthcare industry.

Given the evolution and expansion of secondary use of the ICD-9-CM classification system, its ability to exhibit validity and reliability for its applied metrics is paramount. Many studies were conducted involving ICD-9's sensitivity and predictive ability of PSIs and adverse events shortly after AHRQ's introduction of PSIs and CMS' inclusion of these indicators in quality and safety reporting. Ultimately, findings of these studies were inconclusive concerning the extent of validity and reliability ICD-9 exhibited in identifying particular PSIs and PSI composite score (Tsang, Palmer, Bottle, Majeed, & Aylin, 2011). As result of numerous studies, AHRQ now incorporates other data such as Present of Admission (POA) indicator, to determine if a condition is more likely to be pre-existing or due to potentially preventable care (Agency for Healthcare Research and Quality, 2010). Additionally, AHRQ continually recommends coding changes and clarifications to improve accuracy of ICD-9 code assignment to flag quality indicators. In general, these changes lead to increases in ICD-9's accuracy and positive predictability of most PSIs.

As ICD-9 becomes more reliable in identifying PSIs, additional applications of the improved administrative data can also be more reliable. In 2013, Sadeghi et al investigated ICD-9's validity in identifying postoperative deep vein thrombosis (DVT) and pulmonary embolism (PE), or PSI 12. The study analyzed data and medical records from a sample of total hip and total knee arthroplasty patients from University HealthSystem Consortium (UHC) and AHRQ datasets. Overall, the results showed an 81% positive predictive of ICD-9 codes appropriately identifying PSI 12 from the dataset (Sadeghi et al., 2013). This is an increase of 43-48% from previous studies involving ICD-9 and PSI 12. Implications of these results include reliability of ICD-9 codes when identifying this particular PSI and improvements in administrative data. However, the datasets mostly included patients from large hospitals and academic medical centers. Therefore, coding guidelines and facility coding practices could influence results as previous studies have shown great amount of variance in such practices across health systems (Sadeghi et al., 2013). Given the limited dataset sizes and hospital types, further studies must be conducted to confirm the improvements are due to new coding guidelines and exhibit external validity nationwide.

While this study shows promise for future applications of ICD-9, there have yet to been conclusive results from numerous studies to confirm ICD-9's validity and reliability in identifying PSIs and adverse events. Given the recent implementation of VBP and public hospital profiling, uncertainties remain in ICD-9's ability to adequately represent the quality of care received and measure confounding factors such as illness severity that may affect outcomes and incidence of indicators.

Federal Quality Metric's Bias against Teaching Hospitals

CMS' new Hospital VBP program is a mandated quality program created to reward acute care hospitals with incentive payments based on performance on specific quality metrics. The two routes which facility can qualify for these incentives are through improvement or achievement points. Achievement points are awarded based on an individual hospital's rates during the performance period (i.e. current FY) in comparison to all hospital's rates during the baseline period (Centers for Medicare & Medicaid Services, 2013b). Conversely, comparing the individual hospital's rates during the performance period with the same hospital's rates from the baseline period determines improvement points awarded (Centers for Medicare & Medicaid Services, 2013b). Hospitals will be awarded points based on the better of the two comparisons. Thus, if the magnitude of improvement is greater than the overall achievement score compared to the national benchmark, hospitals receive points proportional to their improvement. The intent of this two-route method is to create an equal and fair opportunity to qualify for points.

For FY 2013, the VBP program incorporates 12 processes of care (POC) measures that carry a 70% weight when determining total performance score for incentive payments. Dupree, Neimeyer & McHugh (2014) break down 7 of these POC measures, primarily surgical measures, to compare differences in surgical performance due to many factors through a multivariate analysis of hospitals from CMS' Hospital Compare dataset. The study broke down the discrepancies in performance based on the following hospital characteristics: size, ownership, teaching status, region, rural location and surgical volume. Overall, higher performance scores were realized in small, private, for-profit and nonteaching hospitals, particularly in non-rural locations regionally located in the Northeast and South (Dupree, Neimeyer, & McHugh, 2014). Additionally, they determined a bias in the "two route method" as most hospitals earned incentive payments via the achievement route. Subsequently, these hospitals received more points overall and had higher composite scores, than those qualifying via improvement (Dupree, Neimeyer, &

McHugh, 2014). Public hospitals mean composite scores were 15.6% lower than private, for-profit hospitals (Dupree, Neimeyer, & McHugh, 2014); therefore, indicating public hospitals were at greater risk for poor performance based on current metrics.

In a similar study, Spaulding, Zhao & Haley (2014) investigated the relationship between HAC scores and total performance scores amongst hospitals. Through analysis of the Hospital Compare datasets, the authors were unable to establish a positive association between high total performance scores and improved quality and safety scores indicated by HAC scores (Spaulding, Zhao, & Haley, 2014). The authors did not isolate individual factors such as patient demographics, volume that could influence total performance scores. Rather, they sought to use a national dataset to investigate trends that could be generalizable over populations (Spaulding, Zhao, & Haley, 2014).

Both of these studies depict potential concerns and biases in new federal quality and performance measures. It should be noted that both studies were conducted using FY 2013 total performance metrics that only include process of care and patient care measures. As of this year, the measures have expanded to include outcome (which PSIs fall under) and efficiency domains. Nonetheless, the implications of these initial reliability studies are profound.

In particular, Dupree, Neimeyer, & McHugh (2014) discover a potential bias against public institutions concerning surgical performance. Quality and performance improvement initiatives can be lengthy and resourcefully burdensome processes for large public institutes. Often, these require a great deal of focus on coding practices and internal data tracking to identify areas and root causes of poor performance (Dupree, Neimeyer, & McHugh, 2014). Many public and teaching institutes may prioritize allocation of such resources to expanding care coverage, education over documentation and tracking. Such institutions are integral to the United States healthcare system; therefore, the burdens and penalties implied by initial VBP metrics could prove detrimental to the cost cutting and quality improving intention of recent legislation (Dupree, Neimeyer, & McHugh, 2014). Trends must be closely monitored as additional measures are included in order to harbor a fair and equal market for hospitals in the US.

AHRQ's PSI-90 Composite Measure

The PSI-90 composite employed in VBP program is slightly different than the PSI-90 composite calculated by AHRQ. VBP's PSI-90 includes only 8 of 11 provider-level PSIs that AHRQ's does, as PSIs 9, 10, and 11 are currently excluded and being monitored, reviewed for inclusion in future fiscal years. CMS publicly releases PSI scores for 4 of 8 individual PSIs (PSI 6, 12, 14 and 15) that make up the composite, as well as the PSI-90. These four publicly released individual measures account for nearly 84% of the PSI-90 composite. The individual PSIs and their respective weights in the calculation of PSI-90 composite are listed in Table 1 below.

Table 1: PSI-90 Composite Measure with Individual Component Measure Weights

Component Measures	*Description*	*Weight in Composite*
PSI-03	Pressure ulcer	0.0226
PSI-06*	Iatrogenic pneumothorax	**0.0708**
PSI-07	Central-line-associated bloodstream infection	0.0652
PSI-08	Postoperative hip fracture	0.0011
PSI-12*	Postoperative pulmonary embolism or deep vein thrombosis	**0.2579**
PSI-13	Postoperative sepsis	0.0742
PSI-14*	Postoperative wound dehiscence	**0.0165**
PSI-15*	Accidental puncture or laceration	**0.4917**
PSI-90*	Composite sum	1.000

*CMS only publicly releases Patient Safety Indicator scores for PSI 6, 12, 14, 15 and 90.

Materials and Methods

Purpose

The purpose of this research is to investigate any bias against teaching hospitals in the AHRQ's current PSI metrics. Given the increased importance of PSI score performance, such a bias could affect reimbursement for teaching hospitals; thus, affecting the delivery of care to patients at teaching hospitals. Recall the following research questions as aforementioned in the objectives section:

1. Are the PSI scores different in teaching versus non-teaching hospitals?
2. Are the patients treated by teaching hospitals different than those treated by non-teaching hospitals as measured by inpatient volume, patient severity and socio-economic status?
3. Is there an interaction between the teaching status of a hospital and the patient demographics that helps explain the difference in PSI scores between teaching and non-teaching hospitals?

Data Source

The Hospital Compare and IPPS Impact File datasets are generated by the Department of Health and Human Services, particularly the Centers for Medicare & Medicaid services, firsthand from claims they receive from accredited providers. Thus, the data is a valid representation of inpatient hospitals and reliable to use in order to generate external validity for the results. Lastly, SPSS Statistics and Microsoft Excel were employed for the descriptive and inferential data.

All time was allocated toward data analysis of the research questions as the FY15 Hospital Compare data has already been collected and released to the public. CMS' Hospital

Compare dataset was used to analyze PSI 6, 12, 14 and 15 scores, as well as the PSI-90 Composite scores across different cohorts.

Data Sample

The CMS compiles data from over 4,000 Medicare-certified hospitals in order to enable the public to compare quality of care—which includes PSI score data—via their Hospital Compare datasets (Centers for Medicare and Medicaid Services, 2015). Furthermore, CMS releases provider claims data for all Medicare-certified acute inpatient hospitals that are reimbursed via CMS' inpatient prospective payments system (IPPS). The Medicare-certified hospital population was separated into teaching and nonteaching hospital samples determined based on Resident-to-Bed Ratio (RBR). This ratio is included in the FY 2015 IPPS Final Rule Impact File dataset (Centers for Medicare and Medicaid Services, 2015). All hospitals having less than 150 hospital beds were excluded from the dataset due to the assumption that smaller hospitals have significant infrastructure differences opposed to their larger counterparts; therefore, inclusion may skew the data analyses. In addition, Maryland hospitals were not included in the analyses as they are exempt from CMS' Hospital VBP Program; thus they are not included in the IPPS data set. Lastly, PSI score data for each hospital from CMS' 2015 Hospital Compare dataset were merged with the IPPS dataset by linking the hospitals through their unique Provider ID number included in both datasets.

Hospital Patient Population Measures

Aside from the PSI score data extrapolated from Hospital Compare dataset, certain patient demographics for each hospital amongst the teaching status cohorts were considered. Specifically, hospitals' inpatient volume, patient severity (or resource intensity) and socio-economic status were analyzed.

- Inpatient Volume- The number of inpatient Medicare claims issued by the hospital will be used as a proxy for inpatient volume.
- Patient Severity- Case-Mix Index (CMI) will be used as a proxy for patients' severity of illness. CMI is based on weights assigned to each Diagnosis-Related Group (DRG) proportional to the resource intensity necessary to treat a patient. The average of these DRG-based weights determines the hospital's CMI.
- Socio-economic status- Disproportionate Share Hospital (DSH) status will be used as a proxy for socio-economic status of each hospital's patient populations. DSH eligibility is determined by the number of Medicaid claims and care delivered to uninsured individuals. CMS decides a hospital is eligible if the hospital DSH patient percentage is greater than 15% of its patient population.

The data used to create the hospital patient population characteristic measures resides in the FY15 IPPS dataset. The variables used for cohort designation and analysis are included below in Table 2.

Table 2: Inclusion/Exclusion Criteria

Variable	Definition	Classification	Type	Source
Provider ID	Unique Medicare provider ID number assigned to each hospital by CMS.	Categorical	Key field for merging data files	Hospital Compare/ IPPS
Teaching Status	If RBR is >0, then Teaching (T). If RBR=0, then Non-teaching (NT).	Categorical	Independent	IPPS
Beds	If bed count <150 =exclude If bed count ≥150 = include	Categorical	Exclusion criteria	IPPS
Patient Safety Indicator (PSI) Score	PSI 6: Iatrogenic pneumothorax PSI 12: Post-op pulmonary embolism or deep vein thrombosis PSI 14: Post-op wound dehiscence PSI 15: Accidental puncture or laceration PSI 90: Composite	Continuous	Dependent	Hospital Compare

Data Analysis

The first research question was addressed by first computing descriptive statistics such as mean, median and standard deviation of PSI scores for each cohort. Analysis for differences in PSI-6, 12, 14, 15 and PSI-90 composite scores across teaching status cohorts were conducted through two-sample t-tests to assess the presence of statistically significant performance differences. A Bonferroni adjusted significance level of 0.01 (0.05/5) will be used to ensure a familywise error rate of 0.05 for the five PSI measures. Additionally, box plots were employed to determine any skewedness (bias) and variance amongst the cohorts.

The second research question was addressed by analyses for differences in patient population demographics— inpatient volume, patient severity and socio-economic status—between teaching and non-teaching hospital cohorts. Descriptive statistics such as mean, median and standard deviation were produced for inpatient volume (bills) and patient severity (CMI) variables. Two-sample t-tests were conducted to determine significant difference between inpatient volume, patient severity and socio-economic status in teaching and non-teaching cohorts. Variables and definitions are listed below in Table 3.

Table 3: Patient Population Demographic Analysis

Variable	Definition	Classification	Type	Source
Teaching Status	If RBR is >0, then T. If RBR=0, then NT.	Categorical	Independent	IPPS
Inpatient Volume	Number of inpatient Medicare claims (bills) by provider	Continuous	Dependent	IPPS
Patient Severity	Hospital specific CMI based on average weight of DRG-assignment	Continuous	Dependent	IPPS
Socio-economic status	Number of Medicaid and uninsured patients (DSH)	Continuous	Dependent	IPPS

The third research question was conditional upon finding statistically significant differences in the first two parts of the analysis. The final question was addressed by first, creating subsets within teaching and non-teaching cohorts for inpatient volume, patient severity and socioeconomic status. Four subsets for each volume, severity and socioeconomic status were created; hospitals were assigned to a subset relative to which quartile their volume, CMI, or DSH percentage resided. Three-Way Multivariate Analysis of Variance (MANOVA) testing were conducted to analyze differences in PSI scores within and between teaching and non-teaching subsets (2), patient volume subsets (4), patient severity subsets (4) and socio-economic subsets (4). Significance levels was set at $p<.05$ for each of these tests. Post-hoc Tukey tests were then conducted to determine which groups' PSI scores are significantly different. Further analyses were then conducted to understand the significance of interaction between patient population variables and teaching status of the hospital and the resulting effect on performance in PSI scores. Variables used for this analysis are listed below in Table 4.

Table 4: PSI Score Analysis Across Teaching Status Subsets

Variable	Definition	Classification	Type
Patient Safety Indicator (PSI) Score	PSI 6: Iatrogenic pneumothorax PSI 12: Post-op pulmonary embolism or deep vein thrombosis PSI 14: Post-op wound dehiscence PSI 15: Accidental puncture or laceration PSI 90: Composite	Continuous	Dependent
Inpatient Volume	Determined by number of inpatient Medicare billing claims per provider: Q1: Small Q3: Large Q2: Medium Q4: Very Large	Categorical	Independent
Patient Severity	Determined by hospital specific CMI score: Q1: Low severity Q3: High Q2: Moderate Q4: Very High	Categorical	Independent
Socioeconomic status	Determined by hospital specific DSH percentage: Q1: Low DSH Q3: High Q2: Moderate Q4: Very High	Categorical	Independent
Teaching Status	If RBR is >0, then T. If RBR=0, then NT	Categorical	Independent

Results

The population of acute-care hospitals that participate in Medicare's IPPS with at least 150 patient beds (n=1,549) was analyzed to determine correlations between teaching status and PSI rates, differences in patient demographics, and interactions between teaching status and patient demographic factors and their ability to account for differences in PSI rates. The population descriptive statistics for PSI scores are presented below in Table 5. Results are organized by individual PSI; number of hospitals in each sample varies based on the number of hospitals with the minimum amount of cases needed to produce a rate for each respective PSI.

Table 5: Descriptive Statistics by PSI for Population

Patient Safety Indicator	N	Mean (SD)	Median
PSI_6_IAT_PTX	1495	0.391 (0.086)	0.370
PSI_12_POSTOP_PULMEMB_DVT	1494	4.362 (1.816)	3.960
PSI_14_POSTOP_DEHIS	1472	1.700 (0.385)	1.580
PSI_15_ACC_LAC	1493	1.810 (0.743)	1.650
PSI_90_SAFETY	1493	0.818 (0.207)	0.770
Valid N (listwise)	*1470*		

Are the PSI scores different in teaching versus non-teaching hospitals?

PSI 6 – Iatrogenic Pneumothorax

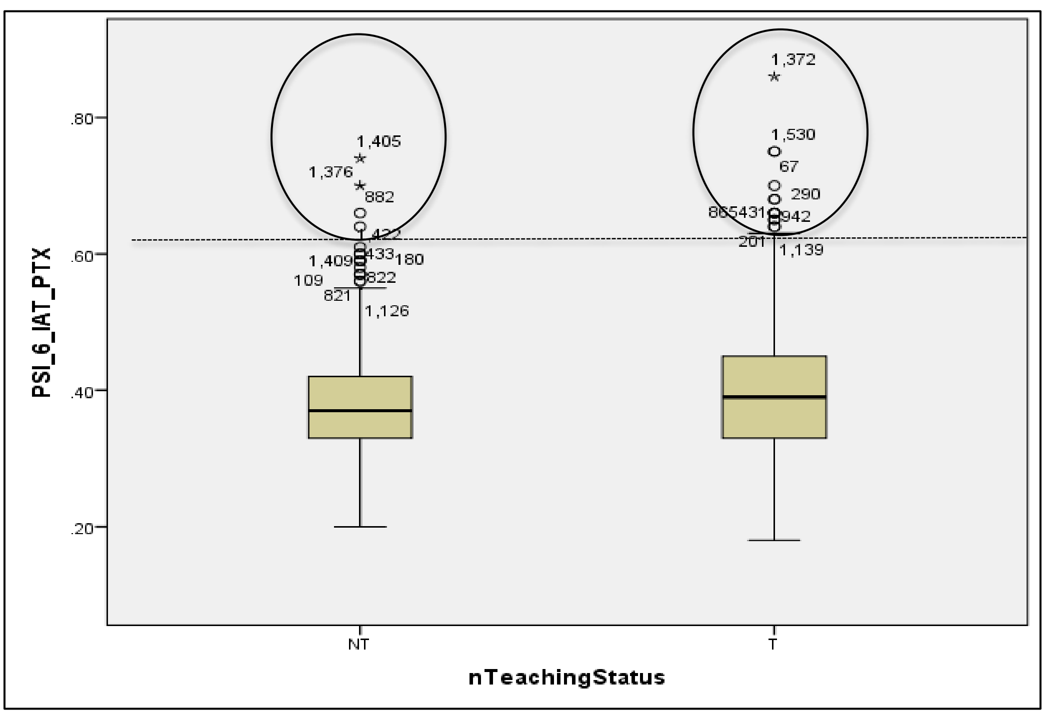
Table 6 summarizes the descriptive statistics for PSI 06 scores by hospital teaching status (n=1495) and the results of the two-sample t-tests between teaching status cohorts. A significant difference in mean PSI 06 scores was found between teaching and non-teaching cohorts ($\rho \leq 0.01$). As indicated by **Figure 1**, the distribution of PSI 6 scores for teaching hospitals is skewed to the right as there is a greater variance and an increased presence of outliers within the teaching subset.

Table 6: Descriptive Statistics, Two-Sample T-Test between Teaching Status Cohorts for PSI 06

Teaching Status	N	Mean (SD)	Median	t-test for Equality of Means (*equal variances assumed)*		
				t	df	p-value
NT	642	0.381 (0.076)	0.360	-4.508	1493	<0.01*
T	853	0.396 (0.092)	0.380			

**Differences are significant at the 0.01 level*

Figure 1: Boxplot of PSI 06 Score Distribution by Teaching Status



PSI 12- Postoperative pulmonary embolism or deep vein thrombosis

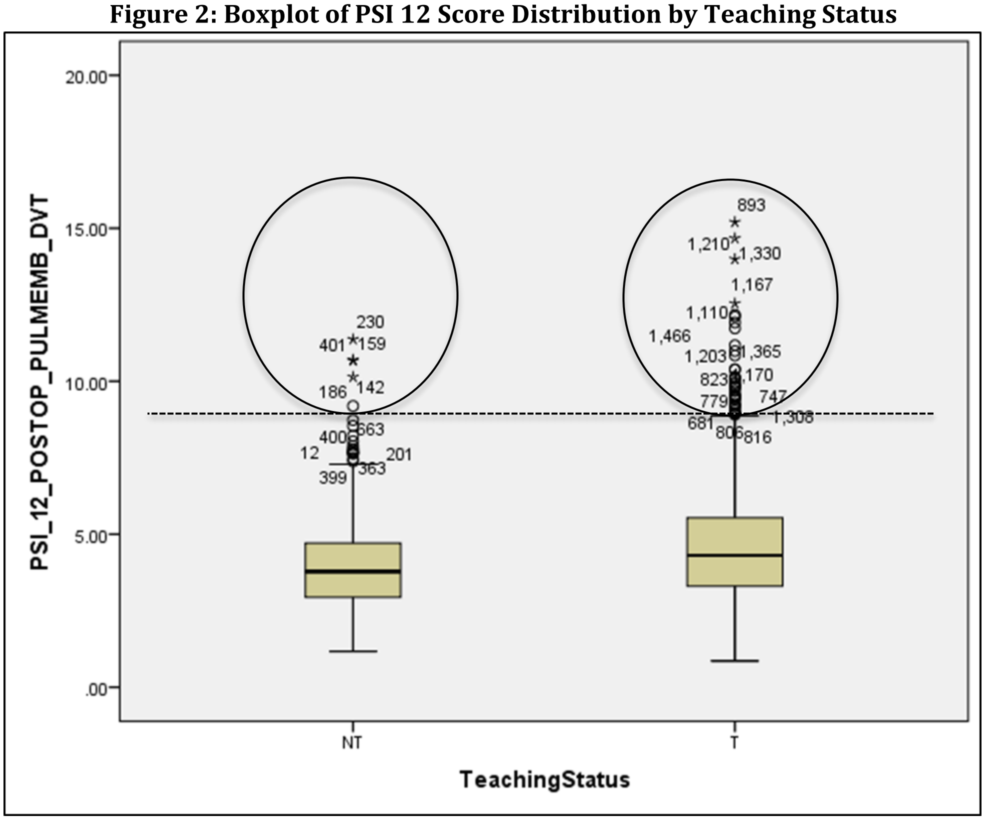
Table 7 summarizes the descriptive statistics for PSI-12 scores by hospital teaching status (n=1494) and the results of the two-sample t-tests between teaching status cohorts. A significant difference in mean PSI 12 scores was found between teaching and non-teaching cohorts ($\rho \leq 0.01$). **Figure 2** depicts a right skew for PSI 12 scores for teaching hospitals as there is a greater variance and an increased presence of outliers within the teaching subset (indicated by circles).

Table 7: Descriptive Statistics, Two-Sample T-Test between Teaching Status Cohorts for PSI 12

Teaching Status	N	Mean (SD)	Median	t-test for Equality of Means (*equal variances assumed)*		
				t	df	p-value
NT	642	3.945 (1.471)	3.690	-7.861	1492	<0.01*
T	852	4.677 (1.982)	4.270			

**Differences are significant at the 0.01 level*

Figure 2: Boxplot of PSI 12 Score Distribution by Teaching Status



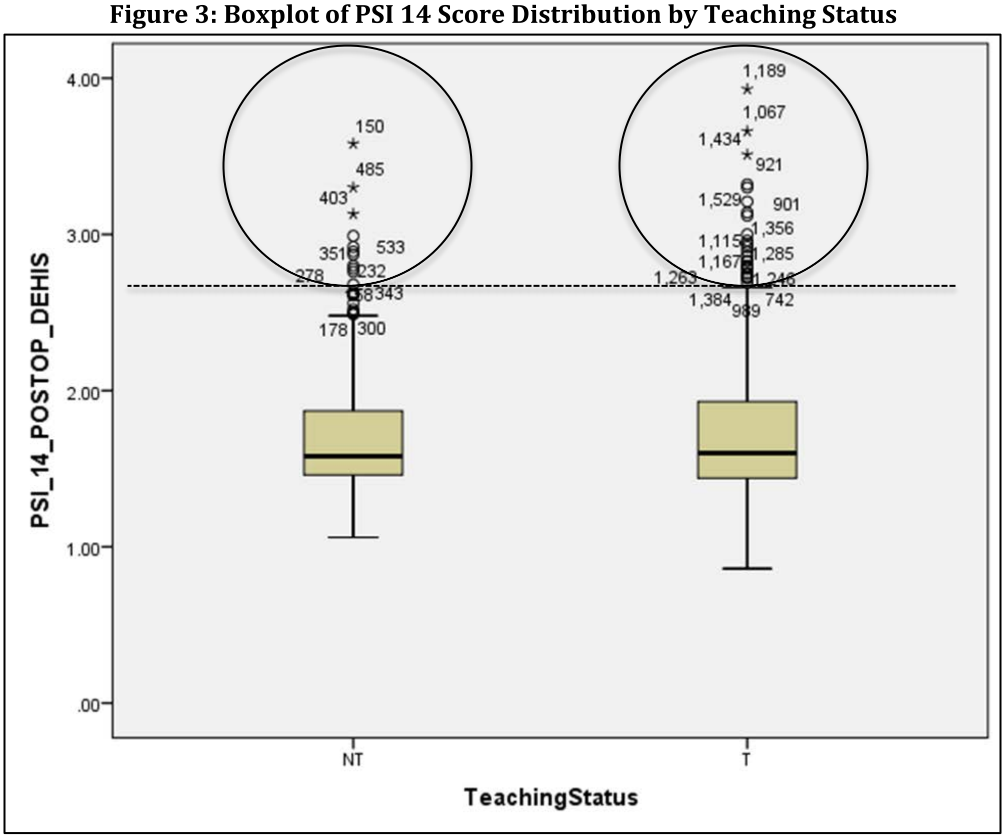
PSI 14- Postoperative Wound Dehiscence

Table 8 summarizes the descriptive statistics for PSI-14 scores by hospital teaching status (n=1472) and the results of the two-sample t-tests between teaching status cohorts. No significant difference in mean PSI 14 score was found between the hospital cohorts. However, **Figure 3** shows that the teaching subset still displays some skewedness to the right in relation to the nonteaching subset. Additionally, it depicts a greater variance and an increased presence of outliers in the teaching subset, yet to a lesser degree in comparison to other PSIs.

Table 8: Descriptive Statistics, Two-Sample T-Test between Teaching Status Cohorts for PSI 14

Teaching Status	N	Mean (SD)	Median	t-test for Equality of Means (*equal variances assumed)*		
				t	df	p-value
NT	637	1.687 (0.350)	1.560	-1.131	1470	0.258
T	835	1.709 (0.410)	1.590			

Figure 3: Boxplot of PSI 14 Score Distribution by Teaching Status



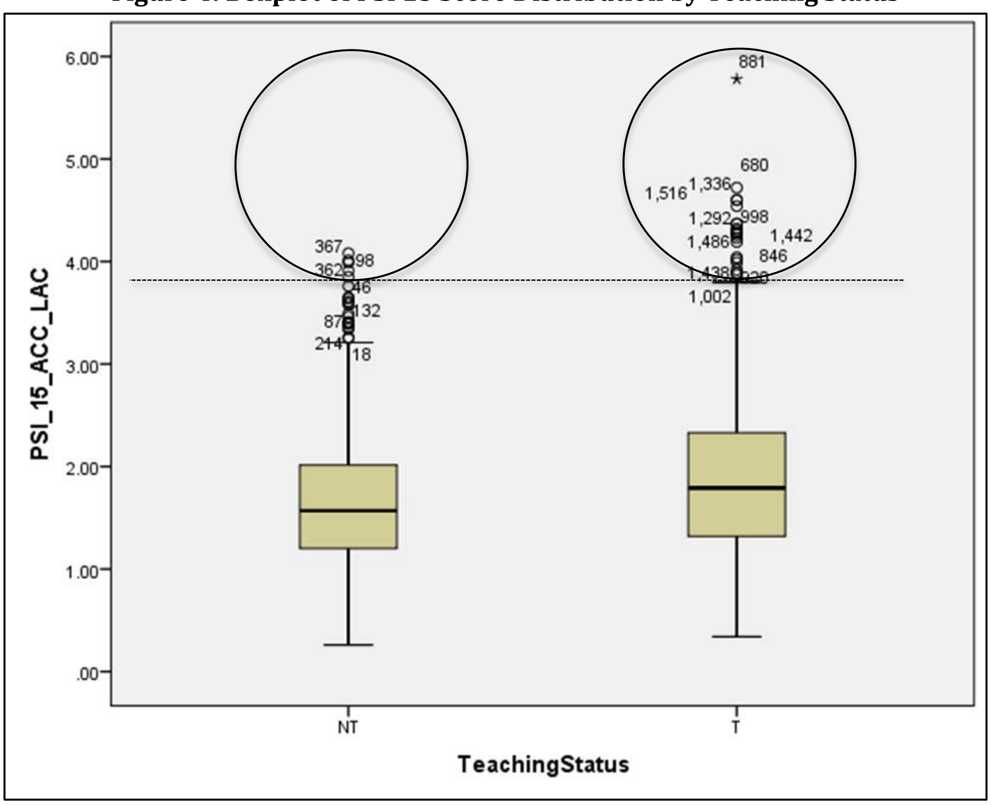
PSI 15- Accidental Puncture or Laceration

Table 9 summarizes the descriptive statistics for PSI-15 scores by hospital teaching status (n=1493) and the results of the two-sample t-tests between teaching status cohorts. A significant difference in mean PSI 15 scores was found between teaching and non-teaching cohorts (ρ≤ 0.01). **Figure 4** depicts a right skew for PSI 15 scores for teaching hospitals as there is a greater variance and an increased presence of outliers within the teaching subset (indicated by circles).

Table 9:- Descriptive Statistics, Two-Sample T-Test between Teaching Status Cohorts for PSI 15

Teaching Status	N	Mean (SD)	Median	t-test for Equality of Means (*equal variances assumed)*		
				t	df	p-value
NT	640	1.678 (0.665)	1.540	-5.929	1491	<0.01*
T	853	1.907 (0.783)	1.770			
**Differences are significant at the 0.01 level*						

Figure 4: Boxplot of PSI 15 Score Distribution by Teaching Status



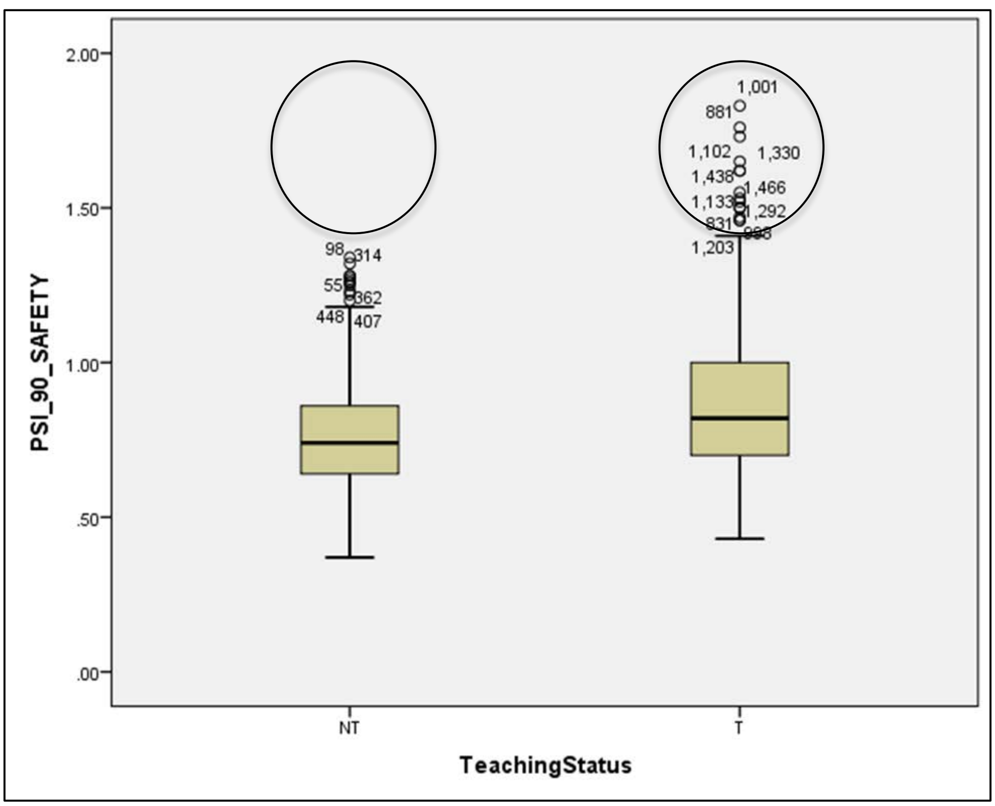
PSI 90- Composite Sum

Table 10 summarizes the descriptive statistics for PSI 90 scores by hospital teaching status (n=1493) and the results of the two-sample t-tests between teaching status cohorts. A significant difference in mean PSI 90 scores was found between teaching and non-teaching cohorts (ρ≤ 0.01). **Figure 5** depicts a right skew for PSI 90 scores for teaching hospitals as there is a greater variance and an increased presence of outliers within the teaching subset (indicated by circles).

Table 10: Descriptive Statistics, Two-Sample T-Test between Teaching Status Cohorts for PSI 90

Teaching Status	N	Mean (SD)	Median	t-test for Equality of Means (*equal variances assumed)*		
				t	df	p-value
NT	640	0.758 (0.172)	0.730	-9.917	1491	<0.01*
T	853	0.863 (0.220)	0.810			
**Differences are significant at the 0.01 level*						

Figure 5: Boxplot of PSI 90 Score Distribution by Teaching Status



Are the patients treated by teaching hospitals different than those treated by non-teaching hospitals as measured by inpatient volume, patient severity and socio-economic status?

Table 11 summarizes the descriptive statistics and statistical test results for inpatient volume (bills), patient severity (CMI) and socio-economic status (DSH) by hospital teaching status. Teaching hospitals have greater mean inpatient volume, patient severity and socioeconomic status in comparison to non-teaching hospitals. Significant differences between hospitals subsets were found (ρ≤ 0.01) for inpatient volume (bills), patient severity (CMI) and socioeconomic status (DSH) variables.

Table 11: Descriptive Statistics, Two-Sample T-Test for Patient Demographic Factors by Teaching Status

Variable	Teaching Status (n)	Mean (SD)	Median (IQR)	t-test for Equality of Means (*equal variances assumed)*		
				t	df	p-value
Bills	NT (672)	4,038 (2,213)	3,626 (2,678)	-8.865	1547	<0.01*
	T (877)	5,445 (3,628)	4,616 (3,893)			
CMI (V33)	NT (672)	1.610 (0.174)	1.607 (0.219)	-12.46	1547	<0.01*
	T (877)	1.751 (0.250)	1.714 (0.296)			
DSH	NT (672)	0.291 (0.157)	0.265 (0.158)	-5.429	1547	<0.01*
	T (877)	0.339 (0.186)	0.299 (0.195)			

**Differences are significant at the 0.01 level*

Is there an interaction between the teaching status of a hospital and the patient demographics that helps explain the difference in PSI scores between teaching and non-teaching hospitals?

Upon finding significant differences in 4 out of the 5 PSIs, inpatient volume, patient severity and socio-economic status variables between teaching status cohorts, analyses were run to better understand the interaction teaching status and patient demographic factors had on differences in PSI scores. In order to do so, each teaching status cohort was split into quartiles based on each respective patient demographic factor.

MANOVA tests revealed significant differences in mean PSI scores between patient severity quartiles were found for PSIs 06, 15 and 90 ($\rho \leq 0.01$). Additionally, significant differences in mean PSI scores between socio-economic status quartiles for PSIs 06, 12, 15 and 90 were found ($\rho \leq 0.05$); however, no significant differences were found between quartiles based on inpatient volume. Upon these discoveries, MANOVA tests were conducted to understand the interactions between teaching status, patient severity and socio-economic status. Significant interactions were found between teaching status and patient severity quartile for PSI 90 ($\rho \leq 0.05$). Meanwhile, significant interactions were found between teaching status and socio-economic status for PSI 06 ($\rho \leq 0.05$). Post-Hoc tests were conducted for those MANOVA tests that were found to be significant to confirm a statistically significant difference in subsets. **Table 12** summarizes the results of the two-sample t-tests between patient demographic factor quartiles and between teaching status and patient demographic quartiles.

Table 12: MANOVA tests within and between Teaching Status and Patient Demographic Factor Subsets

	DSH Quartile			**DSH Quartile * Teaching Status**			**CMI Quartile**			**CMI Quartile* Teaching Status**		
Dependent Variable	df	F	p-value	df	F	p-value	df	F	p-value	df	F	p-value
PSI_06	3	2.89	**<0.05***	3	5.67	**<0.05***	3	5.30	**<0.05***	3	1.51	0.21
PSI_12	3	4.62	**<0.05**	3	1.02	0.38	3	1.19	0.31	3	1.06	0.36
PSI_14	3	2.57	0.053	3	1.35	0.26	3	1.50	0.21	3	0.44	0.72
PSI_15	3	3.19	**<0.05**	3	0.96	0.41	3	15.06	**<0.05***	3	1.99	0.11
PSI_90	3	8.22	**<0.05***	3	1.46	0.23	3	7.61	**<0.05***	3	3.38	**<0.05***

Differences are significant at the 0.05-level
**Post-Hoc tests were found to be significant*

Discussion

Medicare-certified hospitals with 150 beds or more were assigned a teaching status relative to either a zero or non-zero resident-to-bed ratio and then investigated for differences in mean PSI scores, inpatient volume, patient severity, patient socio-economic status and interactions between teaching status and the aforementioned patient demographic factors. The study was conducted using FY 2015 data, and represents the data used for year three of CMS' VBP program. While the respective weight of PSI 90-Composite in determining VBP's total performance score, it will still influence the level of reimbursement experienced by hospitals.

This study first sought to understand the effect teaching status has on mean PSI scores. For all four individual PSIs and composite PSI 90, teaching hospitals had higher mean PSI scores than their non-teaching counterparts; furthermore, differences in three of the four individual PSIs and PSI 90 were found to be statistically significant between teaching status subsets. PSI 14 was the lone indicator where no statistically significant differences were found. The three individual PSI indicators (06, 12 and 15) in which differences were found, comprise of 82% of PSI-90's weighted score. These differences at the individual level account for higher PSI-90 scores which are indicative of poorer performance in CMS' VBP program, and thus, subject to increased payment penalties.

Differences in patient demographic factors between teaching and non-teaching hospitals were then considered. Inpatient volume (bills), patient severity or resource intensity (CMI) and patient socio-economic status (DSH) were greater for teaching hospitals than non-teaching hospitals; differences between teaching status subsets were found to be statistically significant for each factor. So, not only were PSI scores different between teaching and non-teaching subsets but the patient populations in which they treat were different as well. That is, teaching hospitals treated a larger number of patients, sicker or more resource intensive patients, and a larger proportion of their patient populations were poor or uninsured in comparison to non-teaching hospitals.

In order to better understand the effects of patient demographic factors on PSI scores, teaching and non-teaching subsets were split into quartiles based on their respective volume, patient severity and socio-economic status. While no significant differences were found between volume-based quartiles, statistically significant differences were found between socioeconomic status-based quartiles for PSIs 06, 12, 15 and 90. Furthermore, similarly significant differences were found between severity-based quartiles for PSIs 06, 15 and 90. However, significant interactions between teaching status and patient demographic factors were only found between socio-economic quartiles for PSI 06, and between severity quartiles for PSI 90. The latter of the two findings is intriguing, as similar interactions were not found for the individual PSIs that compose composite-90, therefore, differences may lie in the individual measures that CMS does not publicly release.

The results of this study build upon previous studies findings' that question the reliability of federal value-based purchasing metrics in identifying and appropriately reimbursing high quality care. Moreover, it raises additional concerns as PSI-90 is used in additional federal performance programs beyond VBP such as Hospital Acquired Condition (HAC) Reduction program. This introduces redundancy in federal pay-for-performance programs and over penalizing hospitals multiple times for the same metric (Rajaram, Barnard, & Bilimoria, 2015). The methodology in which PSI-90 is calculated and the individual measures that compose it should be re-evaluated for their validity and reliability in identifying and rewarding hospitals that provide quality, safe care to patients.

Conclusion

Ultimately, the results of this study confirm that there exist significant differences in PSI performance for peer groups based on teaching status, patient severity (CMI) and socio-economic status (DSH). These findings indicate one of two conclusions: either teaching hospitals treat sicker patients more poorly than non-teaching hospitals, or the PSI metric does not appropriately account for the underlying risk of an adverse event in the treatment of these patient populations. Regardless, there lie biases in PSI score distribution between teaching and non-teaching hospitals for four out of five PSI measures, including composite PSI-90. Given teaching hospitals higher PSI 90 scores and differences in CMI and DSH factors, these hospitals may be subject to greater penalties in payment for care that is more costly to deliver and less likely to be reimbursed.

Future Directions

This study was limited by its exclusion of hospitals smaller than 150 beds and by its inability to quantify the respective financial impact experienced by individual hospitals relative to their PSI performance. Furthermore, future studies should be conducted to include the other four individual PSI metrics that compose PSI 90 to full understand any biases that exist in their score distribution and their effect on composite PSI 90. Lastly, studies assessing the adequacy of the risk-adjustment methodology employed by CMS for PSI metrics should also be conducted in order to address the discrepancies in patient demographic factors realized in this study.

References

Agency for Healthcare Research and Quality. (2008). Patient Safety Quality Indicators Composite Measure Workgroup.

Agency for Healthcare Research and Quality. (2010). *AHRQ Quality Indicators: Estimating Risk-Adjustment Models Incorporating Data on Present on Admission.*

Agency for Healthcare Research and Quality. (2014, December). AHRQ Quality IndicatorsTM Toolkit for Hospitals | Agency for Healthcare Research & Quality (AHRQ). Retrieved 3 April 2015, from http://www.ahrq.gov/professionals/systems/hospital/qitoolkit/index.html

Alonso, Paula; Love, Tammy R. "Holistic Coding Requires Amped Up Clinical Documentation." *Journal of AHIMA* 84, no.1 (January 2013): 50-51.

Centers for Medicare & Medicaid Services. (2013). *Hospital Value-Based Purchasing Program*.

Centers for Medicare and Medicaid Services. (2015, January 5). CMS.gov. Retrieved 16 July 2015, from http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2015-IPPS-Final-Rule-Home-Page-Items.html

Centers for Medicare and Medicaid Services. (2015, July 16). Data.Medicare.gov. Retrieved 16 July 2015, from https://data.medicare.gov/data/hospital-compare

Dupree, J. M., Neimeyer, J., & McHugh, M. (2014). An Advanced Look at Surgical Performance under Medicare's Hospital-Inpatient Value-Based Purchasing Program: Who is Winning and Who is Losing? *Journal of the American College of Surgeons*, *218*(1), 1–7. doi:10.1016/j.jamcollsurg.2013.09.017

Institute of Medicine. (1999). *To Err is Human: Building a Safer Health System*.

Kaafarani, H. M. A., Borzecki, A. M., & Itani, K. M. F. (2011). Validity of Selected Patient Safety Indicators: Opportunities and Concerns. *Journal of Vascular Surgery*, *54*(5). doi:10.1016/j.jvs.2011.09.074

Kaafarani, H. M. A., & Rosen, A. K. (2009). Using administrative data to identify surgical adverse events: an introduction to the Patient Safety Indicators. *The American Journal of Surgery*, *198*(5), 63–68. doi:10.1016/j.amjsurg.2009.08.008

Naessens, J. M., Campbell, C. R., Shah, N., Berg, B., Lefante, J. J., Williams, A. R., & Culbertson, R. (2011). Effect of Illness Severity and Comorbidity on Patient Safety and Adverse Events. *American Journal of Medical Quality*, *27*(1), 48–57. doi:10.1177/1062860611413456

O'Malley, K. J., Cook, K. F., Price, M. D., Wildes, K. R., Hurdle, J. F., & Ashton, C. M. (2005). Measuring Diagnoses: ICD Code Accuracy. *Health Services Research*, *40*(5p2), 1620–1639. doi:10.1111/j.1475-6773.2005.00444.x

Rahman, M., Neal, D., Fargen, K.M., Hoh, B.L. (2013) Establishing Standard Performance Measures for Adult Stroke Patients: Nationwide Inpatient Sample Database Study. (2013). *World Neurosurgery*, *80*(6), 699-708.

Rajaram, R., Barnard, C., & Bilimoria, K. Y. (2015). Concerns about using the patient safety indicator-90 composite in pay-for-performance programs. JAMA, 313(9), 897. doi:10.1001/jama.2015.52

Sadeghi, B., White, R. H., Maynard, G., Zrelak, P., Strater, A., Hensley, L., … Romano, P. (2013). Improved Coding of Postoperative Deep Vein Thrombosis and Pulmonary Embolism in Administrative Data (AHRQ Patient Safety Indicator 12) After

Introduction of New ICD-9-CM Diagnosis Codes. *Medical Care*. doi:10.1097/mlr.0b013e318287d59e

Spaulding, A., Zhao, M., & Haley, R. D. (2014). Value-based purchasing and hospital acquired conditions: Are we seeing improvement? *Health Policy*, *118*(3), 413–421. doi:10.1016/j.healthpol.2014.10.003

Tsang, C., Palmer, W., Bottle, A., Majeed, A., & Aylin, P. (2011). A Review of Patient Safety Measures Based on Routinely Collected Hospital Data. *American Journal of Medical Quality*, *27*(2), 154–169. doi:10.1177/1062860611414697

Wil, L. (2014). Document Like This, Not That: CDI Insights from the Physician and CDI Specialist Perspective. *Journal of AHIMA*, *85*.